May 1964

ESTIMATION OF A MULTIVARIATE DENSITY*

Theophilos Cacoullos

Technical Report No. 40

University of Minnesota
Minneapolis, Minnesota

*This work was supported in part by the National Science Foundation under Grant Number C-19126.

ERRATA SHEET

for

Estimation of a Multivariate Density

by T. Cacoullos

Page	"Line"	In Place of	Read
3	bottom	$\partial x_1 \quad \partial x_p$	$\partial x_1 \ldots \partial x_p$
4	(1.6)	$K(\frac{x_1-y_1}{h_1}, \ldots)$	$K(\frac{x_1-y_1}{h_1}, \ldots)$
5	(1.12)	$K\frac{y}{h(n)}$	$K(\frac{y}{h(n)})$
7	Table A	$(2\pi)^{p/2}$	$(2\pi)^{-p/2}$
8	middle	$h^{p(1-r)}$	$h^{p(r-1)}$
8		Lemma 1.1	Lemma 2.1
9	middle	$N(\epsilon_1)$	$N(\epsilon_2)$
12	(4.3)	$\rightarrow$	$\sim$
12	(4.4)	$4/p+4$	$1/p+4$
13	7	$\prod_{i=1}^{p} (\frac{e_i}{\sigma_i})$	$\prod_{i=1}^{p} \Phi(\frac{e_i}{\sigma_i})$
19	6	abreviation	observation

ESTIMATION OF A MULTIVARIATE DENSITY

by Theophilos Cacoullos*

University of Minnesota

0. Summary and introduction. The problem of estimating a probability density function has not received as much attention as the corresponding problem of estimating the spectral density of a stationary times series. However, several authors (Rosenblatt [4], Whittle [6], Parzen [3], and Watson and Leadbetter [5]) have recently considered estimating a univariate density function $f(x)$ on the basis of a random sample from $f(x)$.

This paper extends Parzen's results to the case of a multivariate density function. The exposition and most of the results are parallel to those of [3]. Thus a family of asymptotically unbiased, consistent, and asymptotically normal estimates is obtained. Limits for bias and mean square error are also given. Moreover, it is shown that the class of estimators generates an asymptotically Gaussian process with independent components.

The results of the paper, like those of [3], rest rather heavily on Theorem 1.1, which is essentially a modification of some results on approximations of functions at points given in [1]. Actually, the extension is carried out in two directions corresponding to the two general forms of the approximating functions of probability densities as generally described in Theorems 1.1 and 6.1.

Estimating a multivariate density function might arise in various practical situations, e.g., in estimating the hazard function $f(x)/\{1-F(x)\}$.

---

*This work was supported in part by the National Science Foundation under Grant Number C-19126.

Thus, if X_1 and X_2 denote the ages of husband and wife respectively, then $f(x_1, x_2)$ is involved in what actuaries might call "the joint force of mortality" of a couple: $f(x_1,x_2)dx_1dx_2/\{1-F(x_1,x_2)\}$; this is the conditional probability of the husband's death in $(x_1, x_1 + dx_1)$ given survival to age x_1 and the wife's death in $(x_2, x_2 + dx_2)$ given survival to age x_2. Apparently however, such applications at present are rare (if not nonexistent), because of theoretical and technical difficulties involved in such a pursuit.

1. A family of asymptotically unbiased estimates. Let $X_1,X_2,\ldots,X_n$ be n independent observations on a p-dimensional random variable X with absolutely continuous distribution function $F(x)$ so that

$$(1.1) \qquad F(x) = F(x_1,\ldots,x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(y_1,\ldots,y_p)dy_1,\ldots,dy_p \ .$$

We are interested in estimating $f(x)$ on the basis of the sample $X_1,\ldots,X_n$. It should be recalled at the outset that, whereas an unbiased estimate of $F(x)$ is provided by the empirical distribution function

$$(1.2) \qquad F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \epsilon(x - X_i),$$

where $\epsilon(y) = 1$ if $y_i \geq 0$ for all $i = 1,\ldots,p$, and $\epsilon(y) = 0$ otherwise, there exists no uniformly (in x) unbiased estimate of the density $f(x)$ (see [4]). Therefore, it is desirable to look for estimators which, besides having other optimality properties, are unbiased in the limit as n tends to infinity. In the absolutely continuous case a simple asymptotically unbiased estimate may be constructed from $F_n(x)$ as follows.

Let $R(x;h)$ denote the rectangle in E_p centered at x defined as

(1.3) $\quad R(x;h) = \{y: x_i - h_i \leq y_i \leq x_i + h_i,\ i = 1,\ldots,p\}$,

where $h_1,\ldots,h_p$ are positive constants. Then, for "small" rectangles, the density $f(x)$ may be estimated by the "average empirical density function"

$$f_n^o(x) = \left[n \prod_{i=1}^{p} (2h_i)\right]^{-1} \left\{\text{no. of X's falling in } R(x;h)\right\} .$$

This can be written as

$$f_n^o(x) = \left[2^p \prod_{i=1}^{p} h_i\right]^{-1} \Delta F_n(x)$$

where

$$\begin{aligned} \Delta F_n(x) = {} & F_n(x_1 + h_1,\ldots,x_p + h_p) - F_n(x_1 - h_1, x_2 + h_2,\ldots,x_p + h_p) \\ & - \cdots - F_n(x_1 + h_1,\ldots,x_{p-1} + h_{p-1}, x_p - h_p) \\ & + \cdots\cdots + (-1)^p F_n(x_1 - h_1,\ldots,x_p - h_p) \end{aligned}$$

is a p-th order difference of $F_n(x)$. Let us now choose $h = (h_1,\ldots,h_p)$ as a function of the sample size n (this will be understood in the sequel even if, for convenience, we write h instead of $h(n)$) so that

(1.4) $\quad \lim_{n\to\infty} h(n) = 0.$

It follows then, since $F_n(x)$ is an unbiased estimate of $F(x)$ and $F(x)$ is by hypothesis absolutely continuous, that

$$(1.5) \quad \lim_{n\to\infty} E[f_n^o(x)] = \lim_{n\to\infty} \frac{\Delta F(x)}{\prod_{i=1}^{p} (2h_i)} = \frac{\partial^p F(x_1,\ldots,x_p)}{\partial x_1 \cdots \partial x_p} = f(x)$$

whenever x is a continuity point of f.

The estimate $f_n^o(x)$ however may be written also as a weighted average over $F_n(x)$:

$$(1.6)\qquad f_n^o(x) = \left[\prod_{i=1}^{p} h_i\right]^{-1} \int K\left(\frac{x_1 - y_1}{h_i}, \ldots, \frac{x_p - y_p}{h_p}\right) dF_n(y)$$

$$= \left[n \prod_{i=1}^{p} h_i\right]^{-1} \sum_{j=1}^{n} K\left(\frac{x_1 - X_{j1}}{h_i}, \ldots, \frac{x_p - X_{jp}}{h_p}\right),$$

where the weighting function K(y) is the uniform kernel:

$$(1.7)\qquad \begin{aligned} K(y) &= 2^{-p} \quad \text{if } |y_i| \le 1 \text{ for each } i=1,\ldots,p, \\ K(y) &= 0 \quad \text{otherwise;} \end{aligned}$$

here, and henceforth unless otherwise stated, the domain of integration is the entire range of the integrated variable; $X_j = (X_{j1}, \ldots, X_{jp})$, $j=1,\ldots,n$. The form of the estimate $f_n^o(x)$ in (1.6) suggests that by choosing different kernels K(y) as weighting functions we can generate a family of estimates of the form (1.6).

Indeed, we are now going to give fairly general conditions on kernels K(y) so that the corresponding estimates of the form (1.6) are asymptotically unbiased. First, we shall consider the case where the role of the rectangle (1.3) above is played by a square centered at x, so that $h_1=h_2=\ldots=h_p=h$. Therefore, the problem reduces to finding conditions on K(y) under which estimates of the form

$$(1.8)\qquad f_n(x) = \int \frac{1}{(h(n))^p} K\left(\frac{x-y}{h(n)}\right) dF_n(y) = \frac{1}{n(h(n))^p} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h(n)}\right)$$

are asymptotically unbiased in the sense that, whenever the sequence of positive constants $h(n)$ satisfies (1.4), we have

$$\lim_{n \to \infty} E[f_n(x)] = f(x).$$

Such sufficient conditions on K are essentially given by the following theorem, which is a multivariate analog of Theorem 1A of [3], and forms the basis of this paper.

<u>Theorem 1.1.</u> Suppose $K(y)$ is a Borel function on E_p such that

$$(1.9) \qquad \sup_{y \in E_p} |K(y)| < \infty,$$

$$(1.10) \qquad \int |K(y)| dy < \infty ,$$

$$(1.11) \qquad \lim_{|y| \to \infty} |y|^p |K(y)| = 0,$$

where $|y|$ denotes the length of the vector y.

Let $g(y)$ be another scalar function on E_p such that

$$\int |g(y)| dy < \infty ,$$

and define

$$(1.12) \qquad g_n(x) = \frac{1}{(h(n))^p} \int K \frac{y}{h(n)} \; g(x-y)dy,$$

where $\{h(n)\}$ is a sequence of positive constants satisfying (1.4).

Then at every point x of continuity of g

$$(1.13) \qquad \lim_{n \to \infty} g_n(x) = g(x) \int K(y)dy .$$

<u>Proof</u>. Note that

$$g_n(x) - g(x) \int K(y)dy = \frac{1}{h^p} \int \left[g(x-y) - g(x)\right] K\left(\frac{y}{h}\right) dy.$$

Choose $\delta > 0$ and split the region of integration into two regions:

$|y| \leq \delta$ and $|y| > \delta$. We have

$$|g_n(x) - g(x)\int K(y)dy| \leq \max_{|y|\leq\delta}|g(x-y) - g(x)| \int_{|z|\leq\frac{\delta}{h(n)}} |K(z)|dz$$

$$+ \int_{|y|>\delta} \frac{|g(x-y)|}{|y|^p} \frac{|y|^p}{(h(n))^p} \left|K\left(\frac{y}{h(n)}\right)\right|dy + |g(x)| \int_{|y|>\delta} \frac{1}{(h(n))^p} \left|K\left(\frac{y}{h(n)}\right)\right|dy$$

$$\leq \max_{|y|\leq\delta} |g(x-y) - g(x)| \int|K(z)|dz + \frac{1}{\delta^p} \sup_{|z|\geq \frac{\delta}{h(n)}} |z|^p|K(z)| \int |g(y)|dy$$

$$+ |g(x)| \int_{|z|> \frac{\delta}{h(n)}} |K(z)|dz ,$$

which tends to 0 if we let first $n \to \infty (h(n) \to 0)$ and then $\delta \to 0$.

Corollary 1.1. The estimates defined by (1.8) are asymptotically unbiased provided the constants $h(n)$ satisfy (1.4) and the kernel $K(y)$ satisfies, in addition to (1.9) - (1.11), the condition

$$(1.14) \qquad \int K(y)dy = 1.$$

Proof. Note that by (1.8)

$$(1.15) \qquad Ef_n(x) = E\left[\frac{1}{(h(n))^p} K\left(\frac{x - X}{h(n)}\right)\right] = \int \frac{1}{(h(n))^p} K\left(\frac{x - y}{h(n)}\right) f(y)dy$$

and apply Theorem 1 with $g_n(x) = f_n(x)$, $g(x) = f(x)$.

Remarks. It should be observed that $K(y)$ above does not have to be positive in order that Corollary 1.1 hold. However, since we would like $f_n(x)$ to be nonnegative for every x and every n, it is more natural for our purposes to assume that $K(y)$ is also nonnegative, in which case the estimates $f_n(x)$, as well as the $K(y)$, are themselves density functions. We might think, then, of the kernel $K(y)$ as a weighting function, and, moreover, from the definition of $f_n(x)$, for each x as a parameter point K defines an a priori density on E_p of the form $K(\frac{x-y}{h})/h^p$ so that $f_n(x)$ may be considered, in some sense, as a Bayes estimate of $f(x)$ with respect to K as "prior weighting distribution." The form $K(\frac{x-y}{h})$ as a function of y also motivates the symmetry assumption we are going to make in the sequel (c.f. (1.6) and (1.7)), namely, that $K(y)$ is an even function in the sense that

(1.16) $\quad K(y) = K(-y)$

Some examples of $K(y)$ are given in Table A. All the kernels except the normal one are product kernels in the sense that

(1.17) $$K(y) = \prod_{i=1}^{p} K_0(y_i),$$

where $K_0(t)$ is a kernel in E_1. Of course the normal kernel becomes a product kernel if A is a diagonal matrix.

TABLE A

$K(y)$	$k(u) = \int e^{iu'y}K(y)dy$	$\int K^2(y)dy = (2\pi)^{-p}\int k^2(u)du$
2^{-p}, $\lvert y_i\rvert \le 1$, $i=1,\ldots,p$ 0, otherwise	$\prod_{i=1}^{p}\left(\frac{\sin u_i}{u_i}\right)$	$\frac{1}{2^p}$
$(2\pi)^{p/2}\lvert A\rvert^{\frac{1}{2}} e^{-\frac{1}{2}y'Ay}$	$e^{-\frac{1}{2}u'A^{-1}u}$	$\frac{\lvert A\rvert^{\frac{1}{2}}}{(2\sqrt{\pi})^p}$
$2^{-p} e^{-\sum_{i=1}^{p}\lvert x_i\rvert}$	$\prod_{i=1}^{p}(1+u_i^2)^{-1}$	2^{-p}
$\prod_{i=1}^{p}\left[\pi(1+y_i^2)\right]^{-1}$	$e^{-\sum_{i=1}^{p}\lvert u_i\rvert}$	π^{-p}
$\prod_{i=1}^{p}(1-\lvert y_i\rvert)$, $\lvert y_i\rvert \le 1$ 0 otherwise	$\prod_{i=1}^{p}\left(\frac{\sin(u_i/2)}{u_i/2}\right)^2$	$(\frac{2}{3})^p$

2. Asymptotic moments of the estimates. The estimate $f_n(x)$ defined in (1.8) may be written as an average

$$(2.1) \qquad f_n(x) = \frac{1}{n} \sum_{j=1}^{n} \xi_{nj}(x), \; \xi_{nj}(x) = \frac{1}{h^p} K\left(\frac{x-X_j}{h}\right),$$

of independent random variables identically distributed as a random variable

$$(2.2) \qquad \xi_n(x) = K\left(\frac{x-X}{h}\right)/h^p$$

The study of the asymptotic properties of $f_n(x)$ relies on certain asymptotic expressions for the moments of $\xi_n(x)$ given in the following propositions.

Lemma 1.1 Let $r \geq 1$; then, under (1.4), at every continuity point x of f(x)

$$(2.3) \qquad \lim_{n \to \infty} h^{p(r-1)} E\left(\xi_n^r(x)\right) = f(x) \int K^r(y)dy.$$

Proof. By (1.9) and (1.10) $K^r(y)$ is bounded and absolutely integrable, and hence by Theorem 1.1

$$h^{p(1-r)} E\, \xi_n^r(x) = \int \frac{1}{h^p} K^r\left(\frac{x-y}{h}\right) f(y)dy$$

converges to $f(x) \int K^r(y)dy$ as n tends to ∞.

Applying the lemma for $r = 2$ and from (2.1), we obtain

Corollary 2.1. The asymptotic variance of $f_n(x)$ in (1.8) satisfies

$$(2.4) \qquad nh^p \operatorname{Var}[f_n(x)] \xrightarrow{n \to \infty} f(x) \int K^2(y)dy$$

at every continuity point x of f provided the constants $h = h(n)$ satisfy (1.4).

Lemma 2.2. Let x and x^* be two continuity points of f. Then, under (1.4), the asymptotic covariance of $f_n(x)$ and $f_n(x^*)$ satisfies

$$(2.5) \qquad nh^p \operatorname{Cov}\left(f_n(x), f_n(x^*)\right) \longrightarrow 0$$

as $n \to \infty$.

Proof. From (2.1) and (2.2) the quantity in (2.5) is equal to

$$E\left[\frac{1}{h^p} K(\frac{x-X}{h}) K(\frac{x^*-X}{h})\right] - h^p E\left[\frac{1}{h^p} K(\frac{x-X}{h})\right] E\left[\frac{1}{h^p} K(\frac{x^*-X}{h})\right].$$

The second term $\to 0$ as $n \to \infty$ by (1.4) and Theorem 1.1, and the first term after changing the variables can be written as

$$(2.6) \qquad \int K(z)\, K\left(z + \frac{x^*-x}{h}\right) f(x-hz)dz.$$

To show that this also tends to zero as $h \to 0$ (i.e., $n \to \infty$) note that $K(y)$ is bounded by hypothesis and $K(y) \to 0$ as $|y| \to \infty$ by (1.11); hence, it is possible to split the region of integration into two regions, $|z| \leq \rho$ and $|z| > \rho$ where ρ is sufficiently big so that for every $\epsilon_1 > 0$ $K(z) < \epsilon_1$ for $|z| > \rho$, and for every $\epsilon_2 > 0$ there is $N(\epsilon_1)$ such that for all $n > N(\epsilon_2)$ $h(n)$ makes $K\left(z + \frac{x^*-x}{h(n)}\right) < \epsilon_2$ for all $|z| \leq \rho$. Therefore, the integral in (2.6) tends to zero as $h \to 0$ in view of the uniform boundedness of K and the fact that $\int f(y)dy = 1$.

3. Consistency of the estimates.

Theorem 3.1. If the constants $h = h(n)$ in addition to (1.4) satisfy the condition

$$(3.1) \qquad \lim_{n \to \infty} nh^p(n) = \infty,$$

then the estimates $f_n(x)$ are consistent in quadratic mean, i.e.,

$$(3.2) \qquad \lim_{n \to \infty} E[f_n(x) - f(x)]^2 = 0$$

at every point x of continuity of f.

Proof. We have

$$E[f_n(x) - f(x)]^2 = \mathrm{Var}[f_n(x)] + b^2[f_n(x)]$$

where

$$b[f_n(x)] = E[f_n(x)] - f(x)$$

is the bias of $f_n(x)$. Hence and by (3.1) and Corollaries 1.1 and 2.1 (3.2) follows.

Now it will be shown that, under some additional assumptions, $f_n(x)$ is a uniformly consistent estimate of $f(x)$. For this we introduce the Fourier transform

$$(3.3) \qquad k(u) = \int e^{iu'y} K(y)dy$$

of $K(y)$, and the sample characteristic function

$$(3.4) \quad \varphi_n(u) = \int e^{iu'y}\, dF_n(y) = \frac{1}{n}\sum_{j=1}^{n} e^{iu'X_j}$$

Then $f_n(x)$ may be written as

$$(3.5) \qquad f_n(x) = \frac{1}{nh^p}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) = \frac{1}{(2\pi)^p}\int e^{-iu'x}\, k(hu)\, \varphi_n(u)du \text{ ,}$$

since $k(u)$, like $K(y)$, is even.

<u>Theorem 3.2.</u> Uniform consistency of the estimates $f_n(x)$. If

(i) the probability density $f(x)$ is uniformly continuous,

(ii) the constants $h = h(n)$, in addition to (1.4), satisfy

$$(3.6) \qquad \lim_{n\to\infty} nh^{2p} = \infty \text{ ,}$$

(iii) the Fourier transform $k(u)$ of $K(y)$ is integrable (this is true for all $k(u)$ in Table A except the first (c.f.[3])), then, for every $\epsilon > 0$

$$(3.7) \qquad \lim_{n\to\infty} P[\sup_{E_p} |f_n(x) - f(x)| > \epsilon] = 0.$$

<u>Proof</u>. It suffices to show that

$$(3.8) \qquad \lim_{n\to\infty} E^{\frac{1}{2}}[\sup_{E_p} |f_n(x) - f(x)|^2] = 0.$$

But since, by the uniform continuity of f(x) and Theorem 1.1, we have

$$\lim_{n\to\infty} \sup_{E_p} |f_n(x) - f(x)| = 0,$$

(3.8) will hold if

$$(3.9) \quad \lim_{n\to\infty} E^{\frac{1}{2}}[\sup_{E_p} |f_n(x) - E(f_n(x))|^2] = 0$$

holds. By (3.5) we have

$$(3.10) \quad \sup_{E_p} |f_n(x) - E(f_n(x))| \leq (2\pi)^{-p} \int k(hu)|\varphi_n(u) - E(\varphi_n(u))| du,$$

and by Holder-Minkowski inequality (see, e.g., [1])

$$(3.11) \quad E^{\frac{1}{2}}[\sup_{E_p} |f_n(x) - E(f_n(x))|^2] \leq (2\pi)^{-p} \int k(hu)\, E^{\frac{1}{2}}[\varphi_n(u) - E\,{}_n(u)]^2 du$$

$$\leq (n^{\frac{1}{2}}h^p)^{-1} \int k(u)du,$$

which tends to 0 as $n \to \infty$ by (ii) and (iii). (The last inequality in (3.11) follows from the fact that Var $\varphi_n(u) \leq \frac{1}{n}$ by (3.4)).

4. Evaluation of bias and mean square error.

Theorem 4.1. Evaluation of bias. If the probability density function f(x) has continuous partial derivatives of third order in a neighborhood of x and if

$$(4.1) \quad \int y_i \, K(y)dy = 0, \; i = 1,\ldots,p,$$

then as $n \to \infty$

$$(4.2) \quad \frac{b(f_n(x))}{h^2} \to \frac{1}{2} \int d^2f(x;y)\, K(y)dy,$$

where

$$d^2f(x;y) = \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j} y_i y_j.$$

__Proof__. From (1.14) and (1.15)

$$b[f_n(x)] = E[f_n(x)] - f(x) = \int [f(x + hy) - f(x)]\, K(y)dy \ ,$$

and, expanding $f(x + hy)$ by Taylor's Theorem, (4.2) follows in view of (4.1).

Under the assumptions of Theorem 4.1, an approximate expression may be given for the mean square error (m.s.e.):

$$(4.3) \qquad E[f_n(x) - f(x)]^2 \to \frac{f(x)}{nh^p} \int K^2(y)dy + \frac{h^4}{4}\left(\int d^2 f(x;y)\ K(y)dy\right)^2 .$$

The value of h which minimizes the m.s.e. for a fixed value of n is easily found to be (c.f. Lemma 4a of [3])

$$(4.4) \qquad h = \left[\frac{pf(x)\int K^2(y)dy}{n\left(\int d^2 f(x;y)\ K(y)dy\right)^2}\right]^{\frac{4}{p+4}}$$

that is, assuming that the integrals in (4.4) converge absolutely,

$$(4.5) \qquad h = O\left(n^{-\frac{1}{p+4}}\right)$$

Therefore, the m.s.e.

$$E|f_n(x) - f(x)|^2 \sim (p+4)\left[\frac{1}{4p}\left(\int d^2 f(x;y)\ K(y)dy\right)^2\right]^{\frac{p}{p+4}}$$

$$\left(\frac{f(x)}{4n}\int K^2(y)dy\right)^{\frac{4}{p+4}},$$

which shows that $f_n(x)$ as an estimate of $f(x)$ is consistent of order $n^{-\frac{4}{p+4}}$, i.e., its m.s.e. $= O\left(n^{-\frac{4}{p+4}}\right)$ (c.f. [3] and [4] for p = 1).

5. __Asymptotic normality__. In this section, we establish the asymptotic normality of the estimates $f_n(x)$. The proof is based on the expression (2.1) of $f_n(x)$ as a sum of independent and identically distributed random variables $\xi_{nj}(x)$ and the expressions for the asymptotic moments of these

variables as given in section 2.

Theorem 5.1 Let $x^{(1)},x^{(2)},\ldots,x^{(k)}$ be any finite set of continuity points of f. If the constants $h = h(n)$ satisfy (1.4) and (3.1), then the joint distribution of the random variables $f_n(x_1),\ldots,f_n(x_k)$ is asymptotically a k-variate normal in the following sense:

For any real numbers $c_1,\ldots,c_k$,

$$\lim_{n \to \infty} P\left[(nh^p)^{\frac{1}{2}}(f_n(x^{(i)}) - E[f_n(x^{(i)})]) \leq c_i,\ i=1,\ldots,p\right] = \prod_{i=1}^{p} \left(\frac{c_i}{\sigma_i}\right)$$

where Φ denotes the standard normal distribution function, and

(5.1) $\quad \sigma_i^2 = f(x^{(i)}) \int K^2(y)dy,\ i=1,\ldots,p$.

Proof. We have from (2.1) and (2.2)

$$(nh^p)^{\frac{1}{2}}\ f_n(x^{(i)}) = n^{-\frac{1}{2}}h^{p/2} \sum_{j=1}^{n} \xi_{nj}(x^{(i)}),\ i=1,\ldots,k,$$

where, for each fixed i and each n, the

$$\xi_{nj}(x^{(i)}) = h^{-p}K\left(\frac{x^{(i)} - X_j}{h}\right),\ j = 1,\ldots,n$$

are independent random variables identically distributed as a random variable

$$\xi_n(x^{(i)}) = h^{-p}K\left(\frac{x^{(i)} - X}{h}\right).$$

By Bernstein's multivariate central limit theorem (see Robbins-Hoeffding [2]) as applied to the set of independent and identically distributed random vectors

$$Z_{nj} = h^{p/2}\left(\xi_{nj}(x^{(1)}),\ldots,\xi_{nj}(x^{(k)})\right),\ j = 1,\ldots,n\ ,$$

it suffices to show that

$$(5.2) \qquad \lim_{n \to \infty} \operatorname{Cov}\left(h^{p/2}\, \xi_n(x^{(r)}),\ h^{p/2}\, \xi_n(x^{(s)})\right) = \sigma_r^2\, \delta_{rs}$$

for $r,s=1,\ldots,k$, and $\delta_{rs}= 1$ if $r=s$, $\delta_{rs}= 0$ if $r \neq s$, and

$$(5.3) \qquad \lim_{n \to \infty} n^{-\frac{1}{2}}\, \rho_n^3 = 0$$

where

$$\rho_n^3 = \max_{1 \le i \le k} E\left[h^{p/2}\left\{\xi_n(x^{(i)}) - E\left(\xi_n(x^{(i)})\right)\right\}\right]^3 .$$

Now (5.2) follows immediately from (2.3) and Lemma 2.2.

In order to verify (5.3) it is enough to show that

$$(5.4) \qquad n^{-\frac{1}{2}} E\left[h^{p/2} \xi_n(x^{(i)})\right]^3 \longrightarrow 0,\ i=1,\ldots,k,$$

as $n \to \infty$. But by (2.3), for each i, the quantity in (5.4) is approximately equivalent to

$$(nh^p)^{-\frac{1}{2}} f(x^{(i)}) \int K^3(y)dy ,$$

and hence (5.4) follows from (3.1) since $\int K^3(y)dy < \infty$, K being bounded and integrable. This completes the proof of the theorem.

Remark. In view of the results above, we may regard the estimates $f_n(x)$ as a stochastic process with vector parameter x ranging in the domain of definition of the estimated density $f(x)$; furthermore, by Theorem 5.1, the process is asymptotically (as $n \to \infty$) Gaussian and its finite dimensional distributions are multinormal with independent components.

In order to be able to replace $E[f_n(x)]$ by its limit $f(x)$ in Theorem 5.1 so that we can state that $\sqrt{nh^p} f_n(x)$ is asymptotically normal with mean $f(x)$ and variance $f(x) \int K^2(y)dy$, it is necessary to impose some further restrictions on the rate of convergence of h to 0 as a function of n. Thus from Theorem 5.1, the bias of $f_n(x)$ must satisfy

$$\lim_{n \to \infty} (nh^p)^{\frac{1}{2}} b[f_n(x)] = 0$$

which, under the assumptions of Theorem 4.1 and by (3.1), holds if

$$h = 0(n^{-\alpha}), \frac{1}{p+4} < \alpha < \frac{1}{p} .$$

It is interesting however to point out that the above range of α, specifying the rate of convergence of h to 0 as n tends to infinity, does not include the optimum $\alpha^* = (p+4)^{-1}$ corresponding to the h=h(n) of (4.4). Since however α^* is the left end point of the above α interval, it suggests choosing h just smaller than the optimal h. This would make possible the above normal approximation of the distribution of $f_n(x)$ for "large" n, and, in such case, it is clear how this might be used, for example, in setting up a test for the hypothesis that f(x) takes on a specified value. However, the discussion of this and related problems is outside the scope of our present investigation and we will not pursue it any more here.

6. Case of product kernels. The results of the preceding sections depend to some extent on the fact that, roughly speaking, the rectangle R(h,x) in (1.3) was restricted to a square, so that the estimators of the form (1.6) obtained the special form of (1.8). This enabled us to impose fairly general and nice conditions on the weighting functions K(y), which resulted in a natural multivariate generalization as given above. The purpose of this section is to indicate how most of the preceding results carry over to the case of estimates of the form (1.6). Of course, now we assume that the sequence of constant vectors $h(n) = (h_1(n),\ldots,h_p(n))$ satisfies (1.4). Moreover, in order to obtain an approximation theorem analogous to Theorem 1.1, we have to impose a different set of conditions

on the kernels $K(y)$. Such sufficient conditions are given in the following theorem relating to the interesting case of product kernels as defined in (1.17). The estimates now take the form

$$(6.1)\qquad f_n^*(x) = \left[n\prod_{i=1}^{p} h_i\right]^{-1} \sum_{j=1}^{n} \left\{\prod_{i=1}^{p} K_0\left(\frac{x_i - X_{ji}}{h_i}\right)\right\} .$$

<u>Theorem 6.1</u>. Let $K(y)$ be a product kernel in the sense of (1.17), that is,

$$(6.2)\qquad K(y_1,\ldots,y_p) = \prod_{i=1}^{p} K_0(y_i)$$

where K_0 is a kernel in E_1. Suppose K_0 is bounded and absolutely integrable, and

$$(6.3)\qquad \lim_{t\to\infty} |tK_0(t)| = 0 .$$

Let $g(y)$ be as in Theorem 1.1, and define

$$g_n(x) = \frac{1}{\prod_{i=1}^{p} h_i} \int K\left(\frac{y_1}{h_1},\ldots,\frac{y_p}{h_p}\right) g(x-y)dy,$$

where the sequence of constant vectors $h=(h_1,\ldots,h_p)=h(n)$ satisfies (1.4).

Then for every continuity point x of g

$$(6.4)\qquad \lim_{n\to\infty} g_n(x) = g(x)\int K(y)dy.$$

<u>Proof</u>. For the sake of brevity and clarity, we give the proof for the bivariate case $p = 2$, since the general case requires only obvious modifications.

We have by (6.2)

$$(6.5)\qquad g_n(x_1,x_2) - g(x_1,x_2) \iint K_0(y_1)\ K_0(y_2)dy_1dy_2$$

$$= \iint [g(x_1 - y_1, x_1 - y_2) - g(x_1, x_2)] \frac{1}{h_1 h_2} K_0\left(\frac{y_1}{h_1}\right) K_0\left(\frac{y_2}{h_2}\right) dy_1 dy_2 .$$

Let $\delta_1 > 0$, $\delta_2 > 0$ and split the region of integration into four regions,

R_1: $|y_1| \leq \delta_1$ and $|y_2| \leq \delta_2$, R_2: $|y_1| > \delta_1$ and $|y_2| \leq \delta_2$, R_3: $|y_i| \leq \delta_1$ and $|y_2| > \delta_2$, R_4: $|y_1| > \delta$, and $|y_2| > \delta_2$. Then the contribution from R_1 to the absolute value of the right-hand side of (6.5) is not greater than

$$\max_{R_1} |g(x_1 - y_1, x_2 - y_2) - g(x_1, x_2)| \int |K_0(z_1)| dz_1 \int |K_0(z_2)| dz_2 ,$$

which tends to 0 if we let δ_1, δ_2 go to 0, since (x_1, x_2) is a continuity point of g. The contribution from R_2 is at most equal to

$$\frac{1}{\delta_1} \sup_{\substack{|z_1| > \delta_1/h_1 \\ -\infty < z_2 < \infty}} |z_1 K_0(z_1) K_0(z_2)| \iint |g(y_1, y_2)| dy_1 dy_2$$

$$+ \quad |g(x_1, x_2)| \int_{|z_1| > \delta_1/h_1} |K_0(z_1)| dz_1 \int |K_0(z_2)| dz_2 ,$$

which tends to 0 by (6.3) if we let n tend to ∞ (i.e., $h_1(n) \to 0$). Similarly for R_3, and finally the contribution from R_4 does not exceed

$$\frac{1}{\delta_1 \delta_2} \sup_{\substack{|z_1| > \delta_1/h_1 \\ |z_2| > \delta_2/h_2}} |z_1 K_0(z_1)| |z_2 K_0(z_2)| \iint g(y_1, y_2) dy_1 dy_2$$

$$+ \quad |g(x_1, x_2)| \int_{|z_1| > \delta_1/h_1} |K_0(z_1)| dz_1 \int_{|z_2| > \delta_2/h_2} |K_0(z_2)| dz_2 ,$$

which tends to 0 as $n \to \infty$. Hence (6.4) follows.

In obtaining the analogs of Corollaries 1.1, 2.1 and Theorem 3.1, 3.2, and 5.1 for the estimators $f_n^*(x)$ of (6.1) details will be omitted. Thus, for example, the condition (3.1) in Theorem 3.1 will be replaced by $n\prod_{i=1}^{p} h_i(n) \to \infty$ as $n \to \infty$, and (3.6) in Theorem 3.2 by $n\prod_{i=1}^{p} h_i^2(n) \to \infty$ as $n \to \infty$. Theorem 5.1 holds if we replace h^p by $h_1 h_2, \ldots, h_p$; note also that, since $K(y) = \prod_{i=1}^{p} K_0(y_i)$, the asymptotic variances σ_i^2 in (5.1) become $\sigma_i^2 = f(x^{(i)})\left[\int K_0^2(t)dt\right]^p$.

For an estimate of the bias of $f_n^*(x)$ we have the following analog of Theorem 4.1, which can be easily established.

Theorem 6.2. Suppose $f(x)$ satisfies the assumptions of Theorem 4.1, and $\int tK_0(t)dt = 0$. Let $h(n)$ satisfy (1.4) and suppose that

$$\lim_{n \to \infty} \frac{h_i(n)}{h_j(n)} = r_{ij} > 0,\ i \neq j,\ i,j = 1,\ldots,p\ ;$$

then as $n \to \infty$

$$\frac{b[f_n^*(x)]}{|h|^2} \longrightarrow \sum_i \frac{f_{ii}(x)}{r_i^2} \int t^2 K_0(t)dt$$

where

$$f_{ii}(x) = \frac{\partial^2 f(x)}{\partial x_i^2},\ r_i^2 = \sum_{j=1}^{p} r_{ij}^2\ ,\ r_{ii} = 1,\ i=1,\ldots,p\ .$$

Furthermore, it can be easily verified, that for fixed n, the optimum choice of $h(n) = (h_1(n),\ldots,h_p(n))$ in order to minimize the approximate expression for the mean square error of $f_n^*(x)$ (c.f.(4.3)) requires taking $h_1(n) = h_2(n) = \ldots = h_p(n) = h_0(n)$. It then follows that again $h_0(n)$ is of the same order of magnitude as $h(n)$ in (4.4).

Finally, we should like to point out that the estimates $f_n^*(x)$ in (6.1) have a stronger invariance property than the one possessed by the $f_n(x)$ in (1.8), namely, whereas the $f_n(x)$ are invariant under the same scale transformation $X_i \to cX_i (c > 0)$ of each of the components $X_1,\ldots,X_p$ of the abreviation vector X, the $f_n^*(x)$ are invariant under different scale transformations of the components of X, i.e., $X_i \to c_i X_i (c_i > 0)$. This property of $f_n^*(x)$ is more desirable from the practical point of view, since the components of X may represent incommensurable characteristics (e.g., height and weight).

REFERENCES

[1] Bochner, S. (1955). Harmonic Analysis and the Theory of Probability. Univ. of California Press.

[2] Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. Duke Math. J. **15** 773-780.

[3] Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. **33** 1065-1076.

[4] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. Ann. Math. Statist. **27** 832-837.

[5] Watson, G.S. and Leadbetter, M. R. (1963). On the estimation of the probability density, I. Ann. Math. Statist. **34** 480-491.

[6] Whittle, P. (1958). On the smoothing of probability density functions. J. Roy. Statist. Soc., Ser. B **20** 334-343.